PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2001-117920**
(43)Date of publication of application : **27.04.2001**

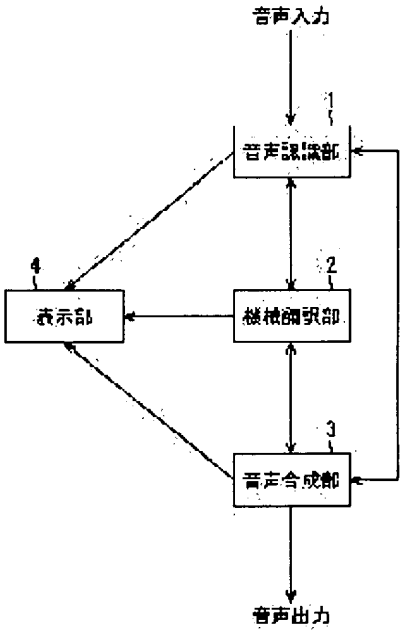
(51)Int.Cl. G06F 17/28
G06F 3/16
G10L 13/00
G10L 15/00
G10L 15/22

(21)Application number : **11-293873**
(22)Date of filing : **15.10.1999**

(71)Applicant : **SONY CORP**
(72)Inventor : **HIROE ATSUO**
MORI HARU
ASANO KOJI
OGAWA HIROAKI

(54) DEVICE AND METHOD FOR TRANSLATION AND RECORDING MEDIUM

(57)Abstract:
PROBLEM TO BE SOLVED: To progressively output a translated sentence, which is provided by a progressive translation, in synthetic voice.
SOLUTION: In a voice recognizing part 1, the voice inputted there is progressively recognized and the recognized result is successively supplied to a machine translation part 2. In the machine translation part 2, the recognized result from the voice recognizing part is progressively translated and a translated sentence is progressively generated. This translated sentence is successively supplied to a voice synthesizing part 3. In the voice synthesizing part 3, this progressively generated translated sentence is compared with the last translated sentence and on the basis of the compared result, reproducing of synthetic voices corresponding to the translated sentence is controlled.



音声翻訳システム

LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] A translation means to translate the input statement by predetermined language, to be translation equipment which outputs the decodement by other language, to translate gradually said input statement inputted gradually, and to generate said decodement gradually, A presentation means to show gradually the decodement gradually generated in said translation means, Translation equipment characterized by having the control means which compares with the last decodement this decodement gradually generated in said translation means, and controls presentation of said decodement by said presentation means based on the comparison result.
[Claim 2] Said control means is translation equipment according to claim 1 characterized by controlling to show the part newly added to said last decodement among said decodements.
[Claim 3] Said control means is translation equipment according to claim 1 characterized by controlling to reshow said decodement when the part corresponding to a modification part in said last decodement is already shown in said presentation means including the part in which said decodement changed said a part of last decodement [at least].
[Claim 4] Said control means is translation equipment according to claim 3 characterized by controlling to reshow said decodement after showing the message of a purport which represents a decodement.
[Claim 5] Translation equipment according to claim 1 which carries out [voice / input] speech recognition gradually, and is characterized by having further a speech recognition means to output the speech recognition result gradually as said input statement.
[Claim 6] Said presentation means is translation equipment according to claim 1 characterized by generating and outputting the composite tone corresponding to said decodement.
[Claim 7] The translation step which translates the input statement by predetermined language, is the translation approach which outputs the decodement by other language, translates gradually said input statement inputted gradually, and generates said decodement gradually, A presentation means to show gradually the decodement gradually generated in said translation step, The translation approach characterized by having the control step which compares with the last decodement this decodement gradually generated in said translation step, and controls presentation of said decodement in said presentation step based on the comparison result.
[Claim 8] The translation processing which translates the input statement by predetermined language and outputs the decodement by other language The translation step which is the record medium with which the program made to perform to a computer is recorded, translates gradually said input statement inputted gradually, and generates said decodement gradually, A presentation means to show gradually the decodement gradually generated in said translation step, The record medium characterized by recording the program equipped with the control step which compares with the last decodement this decodement gradually generated in said translation step, and controls presentation of said decodement in said presentation step based on the comparison result.

[Translation done.]

http://www4.ipdl.jpo.go.jp/cgi-bin/tran_web_cg

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention recognizes gradually the voice especially inputted into translation equipment and the translation approach, and the list about the record medium, for example, translates the speech recognition result gradually, and relates to a record medium at the translation equipment it enables it to show gradually to a user and the translation approach, and a list.
[0002]
[Description of the Prior Art] For example, there is a voice translation system which consists of a voice recognition unit, translation equipment, and a voice synthesizer as a tool for the users which perform the utterance by different language, such as Japanese and English, to aim at communication. In a voice translation system, in a voice recognition unit, speech recognition of the Japanese utterance is carried out, and the speech recognition result is translated into English in translation equipment. And in a voice synthesizer, the translation result is outputted by composite tone. Moreover, in a voice recognition unit, speech recognition of the English utterance is carried out, and the speech recognition result is translated into Japanese in translation equipment. And in a voice synthesizer, the translation result is outputted by composite tone. Therefore, an English speaker (user) can hear a Japanese speaker's utterance in English, and a Japanese speaker can hear an English speaker's utterance in Japanese, and he can have a dialog by understanding a partner's utterance mutually.
[0003] By the way, in the conventional voice translation system, after speech recognition of the whole utterance is performed and the speech recognition result is obtained in a voice recognition unit for example, a translation is performed in translation equipment. And in translation equipment, after the translation of the whole speech recognition result is performed and the translation result is obtained, speech synthesis is performed in a voice synthesizer.
[0004] Therefore, in the conventional voice translation system, after voice was inputted, when time amount until the composite tone corresponding to the translation result of that voice is outputted may have turned into long duration and the so-called between opened for this reason, users' smooth communication might be barred.
[0005] then -- for example, JP,2758851,B and "English-Japanese -- the talk -- gradual generation method" for a language translation, Information Processing Society of Japan, and natural language processing 132-13 There is an approach called the partial translation or gradual translation currently indicated by the reference of 1999.7.23 grade.
[0006] That is, in the gradual translation, in the voice recognition unit, speech recognition is carried out [voice / input] gradually, and, so to speak, the speech recognition result is outputted partially. And in translation equipment, the partial speech recognition result (speech recognition result from the beginning of input voice to current) (suitably henceforth a partial recognition result) from a voice recognition unit is translated gradually, and this outputs the translation result of input voice partially.
[0007]
[Problem(s) to be Solved by the Invention] However, in the gradual translation, it is not indicated by above-mentioned reference about the approach of including in a user the partial translation result (suitably henceforth a partial translation result) of a partial recognition result which translation equipment outputs to a voice output, and showing it gradually to him.
[0008] This invention is made in view of such a situation, and enables it to show gradually the decodement obtained by gradual translation to a user.
[0009]
[Means for Solving the Problem] A translation means for the translation equipment of this invention to translate gradually the input statement inputted gradually, and to generate a decodement gradually, A presentation means to show gradually the decodement gradually generated in the translation means, This decodement gradually generated in the translation means is compared with the last decodement, and it is characterized by having the control means which controls presentation of the decodement by the presentation means based on the comparison result.
[0010] It can be made to control in a control means to show the part newly added to the last decodement among these decodements. Moreover, when the part corresponding to a modification part in the last decodement is already shown in the presentation means including the part in which this decodement changed a part of last decodement [at least], it can be made to control to a control means to reshow this decodement. Furthermore, after showing the message of a purport which represents a decodement, it can be made to control to a control means to reshow this decodement.
[0011] Speech recognition can be gradually carried out [voice / input] to the translation equipment of this invention, and a speech recognition means to output the speech recognition result gradually as an input statement can be further formed in it.
[0012] A presentation means can be made to be able to generate the composite tone corresponding to a decodement, and it can be made to output to it.
[0013] The translation step which the translation approach of this invention translates gradually the input statement inputted gradually, and generates a decodement gradually, A presentation means to show gradually the decodement gradually generated in the translation step, This decodement gradually generated in the translation step is compared with the last decodement, and it is characterized by having

the control step which controls presentation of the decodement in a presentation step based on the comparison result.
[0014] The translation step which the record medium of this invention translates gradually the input statement inputted gradually, and generates a decodement gradually, A presentation means to show gradually the decodement gradually generated in the translation step, This decodement gradually generated in the translation step is compared with the last decodement, and it is characterized by recording the program equipped with the control step which controls presentation of the decodement in a presentation step based on the comparison result.
[0015] The input statement gradually inputted into the translation equipment of this invention and the translation approach, and a list in a record medium is translated gradually, and a decodement is generated gradually. Furthermore, the decodement generated gradually is shown gradually. And this decodement generated gradually is compared with the last decodement, and presentation of a decodement is controlled based on the comparison result.
[0016]
[Embodiment of the Invention] Drawing 1 shows the example of a configuration of the gestalt of 1 operation of the voice translation system (a system means the object with which two or more equipments gathered logically, and it does not ask whether the equipment of each configuration is in the same case) which applied this invention.
[0017] In this voice translation system, for example, if the voice in Japanese is inputted, that voice will be translated and outputted to English, and if the voice in English is inputted, that voice will be translated and outputted to Japanese, and thereby, a Japanese user (speaker) and an English user can have a dialog now.
[0018] That is, the voice which the user uttered is inputted into the speech recognition section 1, and the speech recognition section 1 carries out [voice / which was inputted] speech recognition, and outputs the text as the speech recognition result, and the other accompanying information to the machine translation section 2, a display 4, etc. if needed.
[0019] The machine translation section 2 analyzes the speech recognition result which the speech recognition section 1 outputs, machine-translates the inputted voice with language other than the language of the voice (with the gestalt of this operation, Japanese is translated into English and English is translated into Japanese, respectively), and outputs the text as the translation result, and the other accompanying information to the speech synthesis section 3, a display 4, etc. if needed. The speech synthesis section 3 performs speech synthesis processing based on outputs, such as the speech recognition section 1 and the machine translation section 2, and, thereby, outputs the composite tone as a translation result of other language of the inputted voice etc., for example.
[0020] It consists of liquid crystal displays etc. and a display 4 displays the speech recognition result by the speech recognition section 1 the machine translation result by the machine translation section 2, etc. if needed.
[0021] In the voice translation system constituted as mentioned above, if Japanese voice is inputted, speech recognition of the voice will be carried out in the speech recognition section 1, and it will be supplied to the machine translation section 2, for example. In the machine translation section 2, the speech recognition result by the speech recognition section 1 is machine-translated with English, and is supplied to the speech synthesis section 3. In the speech synthesis section 3, the composite tone corresponding to the translation result by the machine translation section 2 is generated and outputted. Moreover, if English voice is inputted, speech recognition of the voice will be carried out in the speech recognition section 1, and it will be supplied to the machine translation section 2. In the machine translation section 2, the speech recognition result by the speech recognition section 1 is machine-translated with Japanese, and is supplied to the speech synthesis section 3. In the speech synthesis section 3, the composite tone corresponding to the translation result by the machine translation section 2 is generated and outputted.
[0022] Therefore, according to the voice translation system of drawing 1 , he can understand utterance of Japanese by the Japanese user, and a Japanese user can understand utterance of English by the English user, and an English user can have a dialog between a Japanese user and an English user.
[0023] Next, drawing 2 shows the example of a configuration of the speech recognition section 1 of drawing 1 .
[0024] A user's utterance is inputted into a microphone 11 and the utterance is changed into the sound signal as an electrical signal on a microphone 11. This sound signal is supplied to the AD (Analog Digital) transducer 12. In the AD translation section 12, the sound signal which is an analog signal from a microphone 11 is sampled and quantized, and it is changed into the voice data which is a digital signal. This voice data is supplied to the feature-extraction section 13.
[0025] About the voice data from the AD translation section 12, for every suitable frame, the feature-extraction section 13 extracts feature parameters, such as a spectrum, and power, linear predictor coefficients, a cepstrum multiplier, a line spectrum pair, and supplies them to the characteristic quantity buffer 14 and the matching section 15. In the characteristic quantity buffer 14, the feature parameter from the feature-extraction section 13 is stored temporarily. In addition, the feature-extraction section 13 builds in buffer 13A which stores temporarily the voice data which the AD translation section 12 outputs, memorizes the voice data in which the AD translation section 12 carries out a sequential output to buffer 13A, and processes it sequentially.
[0026] The matching section 15 recognizes the voice (input voice) inputted into the microphone 11, referring to the sound model database 16, the dictionary database 17, and the syntax database 18 if needed based on the feature parameter from the characteristic quantity extract section 13, or the feature parameter memorized by the characteristic quantity buffer 14.
[0027] That is, the sound model database 16 has memorized the sound model showing the acoustical descriptions, such as each phoneme in the audio language which carries out speech recognition, and syllable. Here, as a sound model, HMM (Hidden Markov Model) etc. can be used, for example. The dictionary database 17 has memorized the language model which described the chain relation between a word dictionary, a phoneme, or syllable the information about the pronunciation was described to be about each word for recognition (phrase). The syntax database 18 has memorized the syntax rule each word registered into the word dictionary of the dictionary database 17 described it to be how it was carrying out a chain (connected). Here, as syntax rule, a context free language (CFG) and the regulation based on a statistical word chain probability (N-gram) etc. can be used, for example.
[0028] By referring to the word dictionary of the dictionary database 17, the matching section 15 is connecting the sound model memorized by the sound model database 16, and constitutes the sound model (word model) of a word. furthermore, the word model which connected the matching section 15 by referring to the syntax rule memorized by the syntax database 18 in some word models, and was connected by making it such -- using -- a feature parameter -- being based -- for example, HMM -- the voice inputted into the

microphone 11 is recognized by law etc.
[0029] And the speech recognition result by the matching section 15 is outputted in a text etc.
[0030] In addition, when to process again for the inputted voice is required, the matching section 15 processes using the feature parameter memorized by the characteristic quantity buffer 14, and, thereby, needs to require utterance for the second time of a user.
[0031] In the speech recognition section 1, if voice is inputted into a microphone 11, the voice is gradually processed from the time (initiation of the voice section) of being the input initiation. Moreover, by this It is outputted gradually, the recognition result (all the speech recognition result besides some speech recognition results of the inputted voice is also included), i.e., the partial recognition result, of the voice inputted by this time from audio input initiation.
[0032] In the speech recognition section 1, when the voice inputted into the microphone 11 is "accepting **** like this", specifically, the sequential output of a partial recognition result "tea" and "tea" as shown in drawing 3 (A), "tea being drunk", and "drinking tea" is carried out as a sequential input is carried out [voice / the / "**** is accepted like this."]. Moreover, when the voice inputted into the microphone 11 is "the candy is coming down", in the speech recognition section 1, the sequential output of a partial recognition result "a candy" and a "candy" as shown in drawing 3 (B), "it having rained" as the sequential input was carried out [voice / the / "the candy is coming down."], and "it raining" is carried out.
[0033] in addition, the period in a Japanese partial recognition result -- ". -- " -- termination of the voice section is expressed, therefore the phrase as a speech recognition result does not continue after that Moreover, in drawing 3 , the part which the part which attached the underline of a dotted line expressed the part added to the last partial recognition result among these partial recognition results, and attached the underline of a continuous line expresses the part in which the last partial recognition result was changed among these partial recognition results.
[0034] Next, drawing 4 shows the example of a configuration of the machine translation section 2 of drawing 1 .
[0035] The text is analyzed, while the text as a speech recognition result which the speech recognition section 1 outputs etc. is inputted into the text analysis section 21 as an object of machine translation and the text analysis section 21 refers to the dictionary database 24 and the syntax database 25 for analysis.
[0036] That is, the word dictionary in which the notation of each word, part-of-speech information required in order to apply the syntax for analysis, etc. were described by the dictionary database 24 is memorized. Moreover, based on the information on each word described by the word dictionary, the syntax rule for analysis the constraint about a word chain etc. was described to be is memorized by the syntax database 25 for analysis. And based on the word dictionary or the syntax rule for analysis, the text analysis section 21 performs morphological analysis of the text (input text) inputted there, syntax analysis, etc., and extracts language information which constitutes the input text, such as a word and information on functor. Here, there is a thing using the regular grammar, and a context free language and a statistical word chain probability as the analysis approach in the text analysis section 21 etc., for example.
[0037] The language information as an analysis result of the input text obtained in the text analysis section 21 is supplied to the language translation section 22. The language translation section 22 changes the language information on the language of an input text into the language information on the language of a translation result with reference to the language translation database 26.
[0038] That is, the language translation data for changing into the language translation database 26 language information, such as a thesaurus used for count of the bilingual example of the conversion pattern (template) from language information to the language information on output language (language of the output from the language translation section 22) of the source language (language of the input to the language translation section 22), and the source language and output language and the similarity between the bilingual example and source language, are memorized. And in the language translation section 22, the language information on the language of an input text is changed into the language information on output language based on such language translation data.
[0039] The language information on the output language obtained in the language translation section 22 is supplied to the text generation section 23, and the text generation section 23 generates the text which translated the input text into output language from the language information on output language by referring to the dictionary database 27 and the syntax database 28 for generation.
[0040] That is, syntax rule for generation, such as an activity regulation of a word required to memorize the word dictionary in which information, such as a part of speech of a word required to generate the sentence of output language and a conjugated form, was described by the dictionary database 27, and generate the sentence of output language in the syntax database 28 for generation, and constraint of word order, is memorized. And the text generation section 23 changes and outputs the language information from the language translation section 22 to a text based on these word dictionaries and the syntax rule for generation.
[0041] In addition, in the machine translation section 2, as the translation of the input incomplete so to speak from which they are neither a sentence nor a phrase is also to be performed and this mentioned above, sequential translation of the partial recognition result gradually outputted from the speech recognition section 1 is carried out, and it is gradually outputted as a result of [corresponding to it] a translation (i.e., a partial translation result).
[0042] When the voice inputted into the microphone 11 is "accepting **** like this", specifically From the speech recognition section 1, as shown in the same drawing 5 (A) as drawing 3 (A), a partial recognition result "tea", "tea", "tea being drunk", and "tea are drunk. ", although a sequential output is carried out In the machine translation section 2, such a partial recognition result is translated gradually, and thereby, as shown in drawing 5 (B), the sequential output of partial translation result "tea", "tea", "have a cup of tea", and the "I have a cup of tea." is carried out.
[0043] In addition, period"." in a partial translation result expresses termination of a decodement, therefore the phrase as a translation result does not continue after that. Moreover, in drawing 5 , the part which attached the underline of a dotted line expresses the part added to the last partial translation result among these partial translation results.
[0044] So that the partial recognition result of drawing 5 (A) may be compared with the partial translation result of drawing 5 (B) and it may be known Even if a partial recognition result changes from "tea" to "tea", a partial translation result It may not change with "tea". A partial recognition result "for tea to be drunk" from "tea" Having changed slightly is also large from "tea" to "have a cup of tea" ("it drinks" was only added), and a partial translation result may change.
[0045] Next, drawing 6 shows the example of a configuration of the speech synthesis section 3 of drawing 1 .
[0046] The text is analyzed, while the text as a partial translation result which the machine translation section 2 outputs is inputted into

the text analysis section 31 as an object of speech synthesis processing and the text analysis section 31 refers to the dictionary database 34 and the syntax database 35 for analysis.

[0047] That is, syntax rule for analysis, such as constraint about a word chain, is memorized by the dictionary database 34 about the part-of-speech information on each word, and the word which it reads, and the word dictionary in which information, such as an accent, was described is memorized, and was described by the syntax database 35 for analysis at the word dictionary of the dictionary database 34. And the text analysis section 31 analyzes morphological analysis of the text inputted there, syntax analysis, etc. based on this word dictionary and the syntax rule for analysis, and extracts information required for the regulation speech synthesis performed in the latter regulation composition section 32. Here, as information required for regulation speech synthesis, there is phoneme information, such as rhythm information on the information and others for controlling a location, and the accent and intonation of a pause and pronunciation of each word, etc., for example.

[0048] The information acquired in the text analysis section 31 is supplied to the regulation composition section 32, and the voice data (digital data) of the composite tone corresponding to the text inputted into the text analysis section 31 is generated in the regulation composition section 32 using the piece database 36 of a phoneme.

[0049] The piece data of a phoneme are memorized in forms, such as valve flow coefficient (Consonant, Vowel), and VCV, CVC, by the piece database 36 of a phoneme. Namely, the regulation composition section 32 Based on the information from the text analysis section 31, the required piece data of a phoneme are connected and the voice data (voice wave) of the composite tone corresponding to the text inputted into the text analysis section 31 is further generated by adding a pause, an accent, intonation, etc. appropriately.

[0050] This voice data is supplied to the DA translation section 33, and is changed into the sound signal as an analog signal there. This sound signal is supplied to the loudspeaker which is not illustrated, and, thereby, the composite tone corresponding to the text inputted into the text analysis section 31 is outputted.

[0051] In addition, in the speech synthesis section 3, as the voice wave corresponding to the incomplete input used as neither a sentence nor a phrase is also generated and this mentioned above, sequential generation of the voice wave corresponding to the partial translation result gradually outputted from the machine translation section 2 is carried out, and the composite tone corresponding to it is outputted gradually. However, the speech synthesis section 3 compounds a voice wave in which the intonation of the end does not fall, when the partial translation result is completed as a sentence, for example, when compounding a voice wave which lowers the intonation of the sentence end (however, when a sentence is an affirmative sentence) and having not completed as a sentence.

[0052] Moreover, in the speech synthesis section 3, processing (synthetic processing) which generates a voice wave, and processing (regeneration) which carries out D/A conversion of the voice wave, and is outputted as composite tone in the DA translation section 33 are performed to juxtaposition by the regulation composition section 31 if needed. Namely, the regulation speech synthesis section 32 has buffer 32A, the generated voice wave (voice data) can be made to store temporarily at buffer 32A, the DA translation section 33 can read the voice wave memorized by buffer 32A one by one, and synthetic processing and regeneration can be performed now to juxtaposition by regenerating.

[0053] Furthermore, the regulation composition section 32 can recognize now where was reproduced among the voice waves memorized by buffer 32A because the DA translation section 33 has managed the pointer for reading a voice wave from buffer 32A and refers to the pointer. That is, when the voice wave memorized by buffer 32A is "I have a" and the pointer has pointed out "v" of "have" now for example, the regulation composition section 32 can recognize that regeneration of the "v" is performed in the DA translation section 33.

[0054] Next, with reference to drawing 7 , the exchange of the data in each is shown between the speech recognition sections 1 and the machine translation sections 2 in drawing 1 , and between the machine translation section 2 and the speech synthesis section 3. In addition, in drawing 7 , the longitudinal direction shows the exchange of data and the lengthwise direction shows the passage of time.

[0055] If voice is inputted, the speech recognition section 1 will output the message "with voice input" showing that to the machine translation section 2, and will urge preparation of processing to the machine translation section 2. Furthermore, the speech recognition section 1 outputs beginning-of-a-sentence information etc. to the machine translation section 2 if needed. In addition, that the speech recognition section 1 outputs beginning-of-a-sentence information etc. is the case where a beginning-of-a-sentence notation etc. is used in performing speech recognition processing, therefore in not using, the speech recognition section 1 does not output beginning-of-a-sentence information etc.

[0056] If the message "with voice input" from the speech recognition section 1 is received, the machine translation section 2 will output the message "with a translation result" showing the purport which starts the output of a translation result to the speech synthesis section 3, and will urge preparation of processing to the speech synthesis section 3. Furthermore, the machine translation section 2 outputs beginning-of-a-sentence information to the speech synthesis section 3 if needed. In addition, the machine translation section 2 is the case where outputting beginning-of-a-sentence information etc. also uses a beginning-of-a-sentence notation etc. in performing machine translation processing, therefore in not using, the machine translation section 2 does not output beginning-of-a-sentence information etc.

[0057] The speech recognition section 1 outputs a message "with voice input", and further, if the gradual speech recognition to the voice inputted when beginning-of-a-sentence information etc. was outputted if needed is started and a partial recognition result is obtained, it will output the message "with a partial recognition result" showing that to the machine translation section 2. Furthermore, the speech recognition section 1 outputs the obtained partial recognition result to the machine translation section 2 after a message "with a partial recognition result."

[0058] The partial recognition result will be machine-translated and the machine translation section 2 will obtain a corresponding partial translation result, if the partial recognition result which becomes a message "with a partial recognition result", and its message and group is received from the speech recognition section 1. And the machine translation section 2 outputs the message "with a partial translation result" showing the purport from which the partial translation result was obtained to the speech synthesis section 3, and outputs the obtained partial translation result to the speech synthesis section 3 continuously.

[0059] The speech synthesis section 3 will generate and output the composite tone corresponding to the partial translation result, if the partial translation result which becomes a message "with a partial translation result", and its message and group is received from the machine translation section 2.

[0060] Hereafter, the speech-recognition section 1 repeats outputting the group of the partial recognition result and a message "with a

partial recognition result" to the machine-translation section 2, whenever a different partial recognition result (partial recognition result to which the addition of a phrase and deletion are carried out and a change is made to the partial recognition result outputted last time) from the partial recognition result outputted last time is obtained. Similarly, the machine-translation section 2 also repeats outputting the group of the partial translation result and a message "with a partial translation result" to the speech-synthesis section 3, whenever a different partial translation result (partial translation result to which the addition of a phrase and deletion are carried out and a change is made to the partial translation result outputted last time) from the partial translation result outputted last time is obtained. And the speech synthesis section 3 repeats comparing this partial translation result with the last partial translation result, and generating and outputting composite tone from the machine translation section 2, based on the comparison result, whenever it receives a message "with a partial translation result", and a partial translation result.

[0061] As shown in drawing 5 , for example, in addition, this partial recognition result Even if (for example, the 2nd "tea") differs from the last partial recognition result (for example, the 1st "tea") from from [when it can set to drawing 5 (A)], [from / when it can set to drawing 5 (A)] This partial translation result () [<A HREF="/Tokujitu/tjitemdrw.ipdl?N0000=237&N0500=1E_N/;> <>>86=?///&N0001=527&N0552=9&N0553=000007"] 2nd "tea" may not differ from the last partial translation result (from drawing 5 (B) to 1st "tea") from on TARGET="tjitemdrw"> drawing 5 (B). In this case, since a different partial recognition result from that partial recognition result is not obtained after outputting the last partial translation result, the machine translation section 2 does not output a message "with a partial recognition result", and the partial translation result corresponding to this partial recognition result.

[0062] Then, after the voice section is completed (i.e., if the input of the voice to the speech recognition section 1 is completed), the speech recognition section 1 outputs the "voice input termination" message showing that to the machine translation section 2, and outputs to the machine-translation section 2 continuously as a result of [which was finally obtained] partial recognition (i.e., the recognition result of the inputted whole voice) (suitably henceforth a final recognition result).

[0063] The machine-translation section 2 is translating the final recognition result, if a "voice-input termination" message and a final recognition result are received, obtains the translation result (suitably henceforth a final translation result) corresponding to a final partia translation result, i.e., the recognition result of the inputted whole voice, from the speech-recognition section 1, and the "translation termination" message which expresses in termination of a translation, and its final translation result output to the speech-synthesis section 3.

[0064] The speech synthesis section 3 will generate and output the composite tone which lowered the intonation of the sentence end as mentioned above, if a final translation result is received.

[0065] Next, with reference to the flow chart of drawing 8 , actuation of the speech recognition section 1 is explained further.

[0066] First, when it judges with not judging and inputting whether voice was inputted in step S1, the latency time sets the speech recognition section 1 until return and voice are inputted into step S1.

[0067] Moreover, in step S1, when judged with voice having been inputted, it progresses to step S2 and incorporation of the voice is started. That is, the inputted voice is incorporated with a microphone 11 and let it be the voice data as a digital signal by minding the AD translation section 12. This voice data is supplied to the feature-extraction section 13, and sequential storage is carried out at buffer 13A to build in. In addition, incorporation of the above voice is continued as long as an audio input continues.

[0068] As mentioned above, if incorporation of voice is started, it progresses to step S3, and to the machine translation section 2, the speech recognition section 1 will transmit a message "with voice input", and will progress to step S4. In step S4, the beginning-of-a-sentence information about a beginning-of-a-sentence notation etc. is transmitted as opposed to the machine translation section 2 from the speech recognition section 1. In addition, when the beginning-of-a-sentence notation showing the start edge of a text does not exist, in step S4, it is not transmitted at all.

[0069] Then, in step S5, the feature-extraction section 13 reads the voice data memorized by buffer 13A, gives an acoustical treatment to the voice data, and, thereby, extracts a feature parameter. This feature parameter is both supplied to the matching section 15 as if the characteristic quantity buffer 14 is supplied and it memorizes.

[0070] In step S6, the matching section 15 recognizes the voice to this time, after the audio input supplied from the feature-extraction section 13 is started and the input is started using the feature parameter to the voice data to this time, and it obtains a partial recognition result.

[0071] Here, the feature-extraction section 13 will delete the read voice data from buffer 13A, if the voice data memorized by buffer 13A is read. Moreover, the matching section 15 processes using the feature parameter already memorized by the characteristic quantity buffer 14 and the feature parameter newly supplied from the feature-extraction section 13, and thereby, after an audio input is started, it outputs the partial recognition result of the voice to this time.

[0072] After a partial recognition result is obtained, it progresses to step S7 and the matching section 15 judges whether the partial recognition result (this partial recognition result) obtained by processing of this step S6 differs from the partial recognition result (the last partial recognition result) obtained by processing of the last step S6. In step S7, when it judges that it is in agreement with the last partial recognition result, therefore this partial recognition result does not differ, the same processing is repeated by step S5 return and the following.

[0073] In addition, in step S7, if the judgment of whether this partial recognition result differs from the last partial recognition result is performed by comparing the words which constitute each and the word is in agreement, even if the physical relationship (start time and end time of a word) differs, it will be judged with the partial recognition result of this time and last time being in agreement.

[0074] On the other hand, when it judges that this partial recognition result differs from the last partial recognition result in step S7, it progresses to step S8, and to the machine translation section 2, the speech recognition section 1 transmits a message "with a partial recognition result", and progresses to step S9. In step S9, the speech recognition section 1 transmits this partial recognition result to the machine translation section 2, and progresses to step S10.

[0075] At step S10, it is judged whether whether the input of the voice judged that it was inputted at step S1 having been completed, and the voice section were completed. In step S10, when the voice data with which the feature parameter is not extracted is memorized by buffer when judged with the voice section not being completed 13A, the same processing is still repeated by step S5 return and the following at it.

[0076] Moreover, in step S10, when judged with the voice section having been completed (i.e., when voice data is not memorized by buffer 13A), it progresses to step S11, and to the machine translation section 2, the speech recognition section 1 transmits a "voice input termination" message, and progresses to step S12.
[0077] At step S12, the speech recognition section 1 obtains as a result of [of the whole voice section] partial recognition (i.e., the final recognition result of the inputted voice), transmits to the machine translation section 2, and returns to step S1. And it waits to input new voice and the same processing is repeated hereafter.
[0078] Next, with reference to the flow chart of drawing 9 , actuation of the machine translation section 2 is explained further.
[0079] If the latency time is set and a message is first transmitted from the speech recognition section 1 in step S21 until a certain message is transmitted from the speech recognition section 1, the machine translation section 2 will progress to step S22, and will receive the message.
[0080] And in step 23, the class of message from the speech recognition section 1 is judged. In step S23, when it judges that the message from the speech recognition section 1 is a message "with voice input", it progresses to step S24, the translation of the partial recognition result transmitted after that is prepared, and the same processing is repeated by step S21 return and the following. In addition, when beginning-of-a-sentence information etc. is transmitted from the speech recognition section 1 after a message "with voice input" to the machine translation section 2, at step S24, reception of the beginning-of-a-sentence information etc. is also performed.
[0081] Moreover, in step S23, when it judges that the message from the speech recognition section 1 is a message "with a partial recognition result", it progresses to step S25 and the partial recognition result transmitted continuously after that is received.
[0082] And in step S26, the machine translation section 2 translates the partial recognition result from the speech recognition section 1, thereby, obtains the partial translation result as a decodement to the partial recognition result, and progresses to step S27. At step S27, it is judged whether the partial translation result (this partial translation result) obtained by processing of this step S26 differs from the partial translation result (the last partial translation result) obtained by processing of the last step S26. In step S27, when it judges that this partial translation result is in agreement with the last partial translation result, the same processing is repeated by step S21 return and the following.
[0083] Moreover, in step S27, when it judges that this partial translation result differs from the last partial translation result, it progresses to step S28, and to the speech synthesis section 2, the machine translation section 2 transmits a message "with a partial translation result", and progresses to step S29. At step S29, the machine translation section 2 transmits this partial translation result to the speech synthesis section 3, and repeats the same processing return and the following to step S21.
[0084] On the other hand, when it judges that the message from the speech recognition section 1 is a "voice input termination" message in step S23, it progresses to step S30, and from the speech recognition section 1, the machine translation section 2 receives the final recognition result transmitted continuously, and progresses to step S31.
[0085] At step S31, the machine translation section 2 translates the final recognition result from the speech recognition section 1, thereby, obtains the final translation result of the final recognition result, and progresses to step S32. At step S32, the machine translation section 2 transmits a "translation termination" message to the speech synthesis section 3, and progresses to step S33. At step S33, the machine translation section 2 transmits the final translation result obtained at step S31 to the speech synthesis section 3, and repeats the same processing return and the following to step S21.
[0086] Next, with reference to the flow chart of drawing 10 , actuation of the speech synthesis section 3 is explained further.
[0087] If the latency time is set and a message is first transmitted from the machine translation section 2 in step S41 until a certain message is transmitted from the machine translation section 2, the speech synthesis section 3 will progress to step S42, and will receive the message.
[0088] And in step 43, the class of message from the machine translation section 2 is judged. In step S43, when it judges that the message from the machine translation section 2 is a message "with a translation result", it progresses to step S44, generation of the composite tone corresponding to the partial translation result transmitted after that is prepared, and the same processing is repeated by step S41 return and the following. In addition, when beginning-of-a-sentence information etc. is transmitted from the machine translation section 2 after a message "with a translation result" to the speech synthesis section 3, at step S44, reception of the beginning-of-a-sentence information etc. is also performed.
[0089] Moreover, in step S43, when it judges that the message from the machine translation section 2 is a message "with a partial translation result", it progresses to step S45 and the partial translation result transmitted continuously after that is received.
[0090] And in step S46, the speech synthesis section 3 makes the regulation composition section 32 generate the voice wave corresponding to the partial translation result from the machine translation section 2, and progresses to step S49. In addition, about the partial translation result transmitted after a message "with a partial translation result", since a phrase may continue after that, a voice wave in which the intonation of the last of composite tone does not fall is generated.
[0091] On the other hand, when it judges that the message from the machine translation section 2 is a "translation termination" message in step S43, it progresses to step S47, and the final translation result transmitted continuously after that is received, and it progresses to step S48. At step S48, the speech synthesis section 3 makes the regulation composition section 32 generate the voice wave corresponding to the final translation result from the machine translation section 2, and progresses to step S49. In addition, as a result of [which is transmitted after a "translation termination" message] a partial translation (i.e., a final translation result), since a phrase canno continue after that, based on the final translation result, the voice wave in which the intonation of the last of composite tone falls, or the voice wave which goes up is generated. That is, when a translation result with the voice wave final again when a final translation result is a declarative sentence in which the last intonation falls is an interrogative sentence, the voice wave which the last intonation goes up is generated, respectively.
[0092] at step S49, the speech synthesis section 3 compares the voice wave (this voice wave) acquired by processing of this step S46 with the voice wave (the last voice wave) acquired by processing of the last step S46, and, thereby, extracts the voice wave (the following -- suitably -- difference -- it is called a wave) of a different part from the last voice wave among these voice waves.
[0093] namely, -- for example, the case where this voice wave is a thing corresponding to [the last voice wave corresponds to partial translation result "I have a cup of coffee", and] partial translation result "I had a cup of coffee" -- setting -- the part of "had" of these

voice waves at step S49 -- difference -- it is extracted as a wave. in addition, the difference in which the wave from the second half part of "I" in front of "had" to a part for the first portion of "a" just behind "had" differs from the last voice wave among these exact for example, voice waves since a difference arises in fact in this case also into the part (part which is carrying out co-articulation) of the so-called passage in the voice wave of this time and last time -- it is extracted as a wave.

[0094] here -- difference -- if the voice wave corresponding to each partial translation result acquired at step S46 is reproduced as it is, since the same composite tone will be reproduced repeatedly, a wave is searched for for preventing such playback.

[0095] and the difference which it progressed to step S50 and was asked for the speech synthesis section 3 at step S49 -- the composite tone of the part of the last partial translation result corresponding to a wave judges whether it will already be reproduced.

[0096] Namely, for example, the last voice wave is a thing corresponding to partial translation result "I have". [when this voice wave is a thing corresponding to partial translation result "I have a cup of coffee"] Although a difference spectral separation form will serve as a part from the second half part of "have" to the last of "a cup of coffee" among these voice waves if the co-articulation mentioned above is taken into consideration At step S50, it is judged whether the second half part of "have" of the last partial translation result corresponding to the head of the difference wave will already be reproduced.

[0097] In step S50, the composite tone of the part of the last partial translation result corresponding to a difference spectral separation form When judged with not being reproduced yet, up to the location corresponding to the head of a difference wave of the last voice wave When not reproduced yet, it progresses to step S53 and the last playback (output of the composite tone corresponding to the last voice wave) of a voice wave is continued as it is up to the location in front of the difference wave of the last voice wave. And after the last playback of a voice wave which will correspond by just before the location of the head of a difference wave is completed, it progresses to step S54, and from the location corresponding to the head of a difference wave, this playback of a voice wave is started and it progresses to step S55.

[0098] It follows, for example, the last voice wave is a thing corresponding to partial translation result "I have a cup of coffee". [when this voice wave is a thing corresponding to partial translation result "I had a cup of coffee"] Although the wave from the second half par of "I" in front of "had" to a part for the first portion of "a" just behind "had" turns into a difference spectral separation form among these voice wave "I had a cup of coffee" as mentioned above Till just before the second half part of "I" corresponding to the head of the difference wave, the last voice wave "I have a cup of coffee" is reproduced. That is, even a part for the first portion of "I" of the last voice waves is reproduced. And this playback of a voice wave is started after that from the second half part of "I" corresponding to the head of a difference wave of these voice wave "I had a cup of coffee".

[0099] Moreover, for example, the last voice wave is a thing corresponding to partial translation result "I have". [when this voice wave is a thing corresponding to partial translation result "I have a cup of coffee"] Although even the part from the second half part of "have" to the last of "a cup of coffee" serves as a difference spectral separation form among these voice wave "I have a cup of coffee" as mentioned above The last voice wave "I have" is reproduced till just before the second half part of "have" corresponding to the head of the difference wave. That is, even a part for the first portion of "have" of the last voice waves is reproduced. And this playback of a voice wave is started after that from the second half part of "have" corresponding to the head of a difference wave of these voice wave "I have a cup of coffee".

[0100] on the other hand -- step S50 -- setting -- difference -- when it judges that it will already be reproduced by the composite tone of the part of the last partial translation result corresponding to a wave (i.e., when playback after the location corresponding to the head of a difference wave of the last voice wave has already been performed), it progresses to step S51 and the speech synthesis section 3 performs correction processing.

[0101] Namely, for example, the last voice wave is a thing corresponding to partial translation result "I have a cup of coffee". [when this voice wave is a thing corresponding to partial translation result "I had a cup of coffee"] Although the wave from the second half part of "I" of these voice wave "Ihad a cup of coffee" and in front of "had" to a part for the first portion of "a" just behind "had" turns into a difference spectral separation form as mentioned above When the last voice wave "I have a cup of coffee" after the second half part of "I" corresponding to the head of the difference wave is already reproduced That is, for example, when even "have" is reproduced, it is necessary to restate the "have" to "had" in this voice wave.

[0102] Moreover, when the last voice wave corresponds to partial translation result "have a cup of coffee", this voice wave is a thing corresponding to partial translation result "I have a cup of coffee" and "have" of the last voice wave or subsequent ones is already reproduced for example, it is necessary to restate the "have" after "I have" in this voice wave.

[0103] In correction processing, processing for outputting the composite tone of such a correction in a natural form is performed.

[0104] Specifically, it is reproduced as it is to the good location of so to speak the ends of the last voice wave, such as a break of a word or a phrase, and a location of a pause. Then, the composite tone for the correction for tying to a correction automatically is reproduced. namely, -- the case where composite tone is Japanese -- for example, "****" -- "-- not but -- " -- the every day of "it mistook", "****", etc., etc. -- conversation -- saying -- a mistake -- hesitating to say -- etc. -- when it carries out, the composite tone same with speaking is reproduced. Similarly, when composite tone is English, composite tone, such as "woops", and "well", "I mean", is reproduced.

[0105] After that, it progresses to step S52, and this voice wave is reproduced and it progresses to step S55.

[0106] In addition, a pause is left for a while and you may make it reproduce this voice wave in step S52 after that in the correction processing in step S51, without reproducing the composite tone for a correction.

[0107] Moreover, the starting position when reproducing this voice wave in step S52 after whether the composite tone for what kind of correction is reproduced and the correction processing in step S51 can be made to choose by how much parts which should be restated, for example among the already reproduced composite tone there are.

[0108] That is, when the part which should be restated from one word reproduced immediately before when composite tone was Japanese in case what is necessary is just to perform a correction is short, "****" comparatively short as composite tone for a correction is reproduced, and this voice wave can be reproduced from the location corresponding to the head of a difference wave after that. Moreover, conversely, it reproduces comparatively long as composite tone for a correction "it having mistaken", when the part which should be restated was long, and, for example, this voice wave can be reproduced from that head after that (this playback will be restated from the beginning).

[0109] Even the case where the case where the composite tone of the part of the last partial translation result corresponding to a difference spectral separation form is not reproduced yet here is not reproduced yet up to the location corresponding to a difference spectral separation form although the last voice wave is already reproduced, and the last voice wave may not be reproduced yet. Although already reproduced, the last voice wave When not reproduced yet up to the location corresponding to a difference spectral separation form it mentioned above -- as -- step S -- although it progresses to 53 and 54 one by one, when even the last voice wave is not reproduced yet, the last voice wave is canceled, step S51 is skipped, and it progresses to step S52, and it progresses to step S55, this voice wave being used as a reproductive object. That is, to the last voice wave of buffer 32A, this voice wave is overwritten and playback is performed from the location which the above-mentioned pointer has pointed out so that a head may be in agreement.
[0110] At step S55, when judged with it being judged whether the message received at step S42 is a "translation termination" message, and not being a "translation termination" message, the same processing is repeated by step S41 return and the following. Moreover, in step S55, when it judges that a message is a "translation termination" message, the post process of clear and others of buffer 32A which the regulation composition section 32 builds in is performed, and the same processing is repeated by step S41 return and the following.
[0111] In addition, in the machine translation section 2, in performing the above gradual translations, in order to lessen a correction, translating is desirable so that the word order of the language before a translation may not differ from the word order of the language after a translation if possible.
[0112] Such a translation can perform what detected a certain keyword and divided the speech recognition result by the keyword from the speech recognition result translated in the machine translation section 2 by considering as the object (translation unit) of one translation.
[0113] That is, for example, when translating Japanese into English, the conjunctive particle of "-", "that of -", etc., etc., the part used as a termination form, an adonominal modification part, "that of -", "that of -", "-", "-", "- (carry out) etc.", etc. can be made into a keyword A translation not different if possible of word order is attained by setting the part which divided the Japanese text by such keyword and was obtained as the object of a gradual translation.
[0114] here, the part used as a termination form can be set "for the pan to have been eaten, for cow's milk to have been drunk and for it to have left" in the continuous use form of a conjugated word, -- "-- eating -- " -- it means the "drinker" and "it having left". [which connected the sentence] Moreover, when an adonominal modification part means the part to which the noun follows the continuous use form of a conjugated word like "the book just bought yesterday" and translates it into English, it is a part which serves as description using a relative pronoun etc.
[0115] For example, when performing a gradual translation now for the speech recognition result of the voice "it is said that it is difficult to translate Japanese into English, with word order maintained since word order differs greatly." by the Japanese sentence, the machine translation section 2 detects a keyword from the partial recognition result in which the speech recognition section 1 carries out a sequential output.
[0116] here -- "-- word order -- large -- differing -- [-- that --] -- word order -- having maintained -- [-- as --] -- Japanese -- English -- translating -- [-- that --] -- being difficult -- ** -- saying -- having -- **** . -- " -- from -- [--] -- surrounding -- a phrase -- a keyword -- ****** -- detecting -- having -- a thing -- ** -- carrying out .
[0117] In this case, the keyword which the machine translation section 2 detected, "since word order differs greatly it is said that it is difficult to translate Japanese into English, with word order maintained. " -- "translating Japanese into English", "with word order" maintained, and for example, "it being difficult", "since word order differs greatly" -- and -- "-- ** -- it says. " -- it divides, and considers as a translation unit and each translation unit is translated.
[0118] "Since word order differs greatly" now to "Because of largely different word order" "As maintained word order" to "with keeping the word order" It says. ""translating Japanese into English" -- "to translate Japanese to English" -- it is difficult" -- "is difficult" -- "-- ** -- " -- as a final translation result, supposing it translates into "and they say.", respectively "Because oflargely different word order, with keeping the word order, to translateJapanese to English, and they say." are obtained. Therefore, the case of Japanese before a translation and the translation result from which word order does not change so much can be obtained.
[0119] The above processing process is shown in drawing 11 . In addition, drawing 11 (A) shows a partial recognition result, and drawing 11 (B) shows the partial translation result, respectively. Moreover, in drawing 11 (also setting to drawing 12 mentioned later the same), the part which has attached the duplex underline expresses an added part to the last partial recognition result and the last partial translation result.
[0120] here -- the translation unit of the Japanese last -- "-- ** -- it says -- " -- when translating into English, it is common to insert "It is said that" etc. in the beginning of a sentence of a translation result, but if a gradual translation is performed, and it performs such insertion in carrying out the sequential output of the partial translation result obtained as a result, possibility that a correction must be performed from the beginning will become high. So, since the original Japanese sentence "word order differs greatly, it is said here that it is difficult to translate Japanese into English, with word order maintained," "since word order differs greatly it is difficult to translate Japanese into English, with word order maintained. " -- corresponding English -- with "Because of largely different word order, with keeping the wordorder, and to translate Japanese to English" It says. "--" -- corresponding English -- by translating into the form divided into "they say.", it prevents that the above corrections are performed.
[0121] In addition, it becomes possible by performing a Japanese sentence "it having been called -." and the translation with "I think -", same "it being -", etc. to prevent a correction.
[0122] In addition, it is not ["since word order differs greatly, it is said that it is difficult to translate Japanese into English, with word order maintained", and] a gradual translation. After there is an input of the whole, when it translates Generally "Since there is large difference of word order between Japanese and When it comes to English and it is said to be difficult to translate Japanese into English with keeping the word order.", think, but Such a translation result can be made to display by the display 4 while it outputs the translation result by gradual translation by composite tone. Moreover, such a translation result can also be made to supply the speech synthesis section 3 in the machine translation section 2 as a final translation result outputted with the "translation termination" message explained by drawing 7 .
[0123] Next, if it divides by the keyword and divides into a translation unit for example, as mentioned ["since word order differs greatly

in Japanese and English, it is said that it is difficult to translate with word order maintained", and] above for example, -- "-- Japanese -- English -- word order -- large -- differing -- [-- that --] -- " -- "-- word order -- having maintained -- [-- as --] -- " -- "-- translating -- [-- that --] -- " -- "-- being difficult -- " -- "-- [--] -- saying -- having -- **** . -- It becomes " etc. In addition, the part enclosed with [] expresses a keyword.

[0124] In this case, according to the procedure mentioned above, in the machine translation section 2, first, "since word order differs greatly in Japanese and English", the first translation unit is translated into "Because English and Japanese are largely different in word order", and this is outputted as a partial translation result.

[0125] And next, a translation unit "have maintained word order" is translated into "with keeping the word order", and "Because English and Japanese are largely different in word order and with keeping the word order" which added the translation result to the last partial translation result are outputted as a partial translation result.

[0126] Then, it moves to the translation of the following unit "translate." Temporarily, if a partial translation is performed only in this unit, it will translate into "translation is", for example. The translation result The last partial translation result"Because English and Japanese are largely different inword order, with keeping the word By adding to order and ", "BecauseEnglish andJapanese are largely different in word order and with keepingthe word order and translation is" are obtained.

[0127] However, the translation result tends to become unnatural if a partial translation is performed to a shorter translation unit like "translating." For example, it is indefinite whether it is that the direction of a translation result has required "with keeping the word order" for "translation" to a thing with clear "having maintained word order" having started "it translates" in the Japanese sentence of an input in the example given here.

[0128] Then, with its simple substance, when a translation unit is shorter like "translating", after combining with the translation unit before it and generating a new translation unit, without translating, a partial translation is performed. That is, "with keeping the word order" which it is as a result of a partial translation is once canceled "as word order was maintained", in the case of this example, it combines "having maintained word order for as which are newly "translating" and a translation unit in front of that", it generates the translation unit of "translating, with word order maintained", and performs a partial translation to this translation unit. "translation with keeping the wordorder is" is obtained as a result (if it translates in this way, it will become clear that "with keeping the wordorder" has started "translation").

[0129] Thus, according to a partial recognition result being generated gradually, a part of partial translation result may change. However the range which change produces is restricted in a translation unit, and the whole translation result does not change a lot.

[0130] The processing process so far is shown in drawing 12 . In addition, drawing 12 (A) shows a partial recognition result, and drawing 12 (B) shows the partial translation result, respectively. Moreover, in drawing 12 , the part which has attached the single underline expresses the modification part to the last partial recognition result and the last partial translation result. Although the 3rd partial recognition result and partial translation result are bundled with the parenthesis from the top, this is because the 3rd partial translation result is temporary and is not generated in fact. When it is judged that "translating" is included in a partial recognition result and it is short as a translation unit A partial translation result a part Cancellation ("with keepingthe word order" is canceled), Reconstruction (the translation unit of "translating, with word order maintained" is generated) of a translation unit, and rerun ("translationwith keeping theword order is" is generated) of a partial translation are performed, and it will be in the 4th condition of drawing 12 as a result.

[0131] Thus, when change arises in the middle of a partial translation result, according to where [of a partial translation result] speech synthesis is reproducing, it investigates whether there is any need for a correction. When speech synthesis has not reached yet difference (part which change produced), there is no need for a correction (for example, when per ["different"] is being reproduced). On the other hand, since there is the need for a correction when speech synthesis has already reached difference (for example, when per ["keeping"] is being reproduced), speech synthesis is interrupted once. and the above-mentioned "voice for a correction" is outputted, and speech synthesis is resumed after that from the point which change produced from "translation" of a top to the 4th partial translation result the example of drawing 12 .

[0132] the translation after "translating", i.e., the translation by "it is said that it is difficult", -- drawing 11 It is completely the same as that of an example.

[0133] As mentioned above, since this partial translation result generated gradually is compared with the last partial translation result and playback of composite tone was controlled based on the comparison result, the decodement obtained by gradual translation can be gradually shown to a user, and, thereby, time amount until the translation result is shown from voice input can be shortened. Furthermore, since he can understand a partner's utterance in a short time after there is the utterance if it carries out from a user, the response to a partner's utterance also becomes possible [carrying out immediately], after the utterance is carried out. Therefore, it becomes possible through a voice translation system to aim at communication still more smoothly.

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram showing the example of a configuration of the gestalt of 1 operation of the voice translation system which applied this invention.
[Drawing 2] It is the block diagram showing the example of a configuration of the speech recognition section 1.
[Drawing 3] It is drawing for explaining the gradual speech recognition by the speech recognition section 1.
[Drawing 4] It is the block diagram showing the example of a configuration of the machine translation section 2.
[Drawing 5] It is drawing for explaining the gradual machine translation by the machine translation section 2.
[Drawing 6] It is the block diagram showing the example of a configuration of the speech synthesis section 3.
[Drawing 7] It is drawing for explaining the exchange between the speech recognition section 1 and the machine translation section 2 and between the machine translation section 2 and the speech synthesis section 3.
[Drawing 8] It is a flow chart for explaining actuation of the speech recognition section 1.
[Drawing 9] It is a flow chart for explaining actuation of the machine translation section 2.
[Drawing 10] It is a flow chart for explaining actuation of the speech synthesis section 3.
[Drawing 11] It is drawing showing a partial recognition result and a partial translation result.
[Drawing 12] It is drawing showing a partial recognition result and a partial translation result.
[Drawing 13] It is the block diagram showing the example of a configuration of the gestalt of 1 operation of the computer which applied this invention.
[Description of Notations]
1 Speech Recognition Section 2 Machine Translation Section 3 Speech Synthesis Section 4 Display 11 Microphone (Microphone), 12 AD translation section 13 The feature-extraction section and 13A buffer 14 A characteristic quantity buffer, 15 Matching section 16 sound model database, 17 Dictionary database 18 Syntax database 21 Text analysis section 22 Language translation section 23 The text generation section, 24 Dictionary database 25 The syntax database for analysis, 26 Language translation database 27 dictionary database 28 The syntax database for generation, 31 Text analysis section 32 The regulation composition section, 32A Buffer 33 DA translation sections, 34 Dictionary database 35 The syntax database for analysis, 36 Piece database of a phoneme 101 buses 102 CPU 103 ROM 104 RAM 105 A hard disk, 106 Output section The 107 input sections 108 Communications department 109 Drive 110 Input/output interface 111 Removable record medium

[Translation done.]

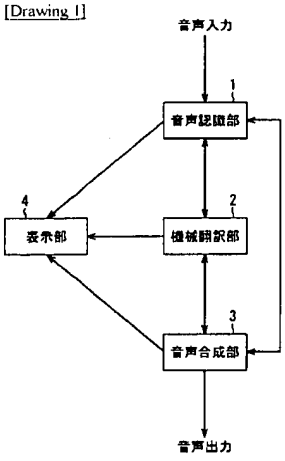
* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

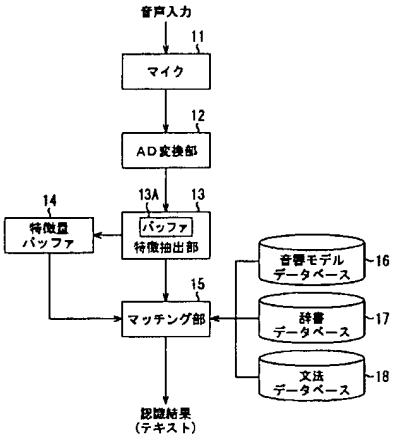
DRAWINGS

[Drawing 1]



音声翻訳システム

[Drawing 2]



音声認識部 1

[Drawing 3]

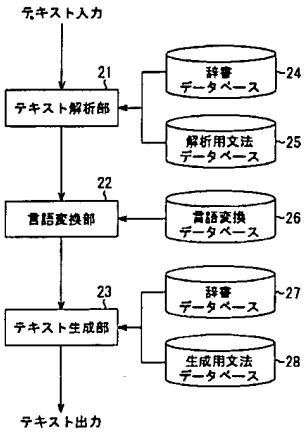
(A)

「紅茶」→「紅茶を」→「紅茶を飲む」→「紅茶を飲む。」

(B)

「飴」→「飴が」→「雨が降った」→「雨が降っている。」

[Drawing 4]

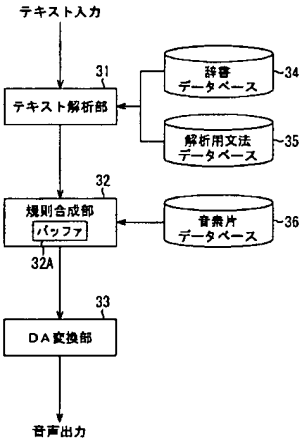


機械翻訳部 2

[Drawing 5]

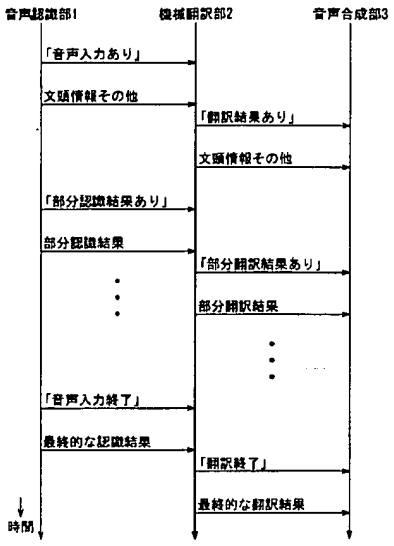
「紅茶」	"tea"
↓	↓
「紅茶を」	"tea"
↓	↓
「紅茶を飲む」	"have a cup of tea"
↓	↓
「紅茶を飲む。」	"I have a cup of tea."
(A)	(B)

[Drawing 6]

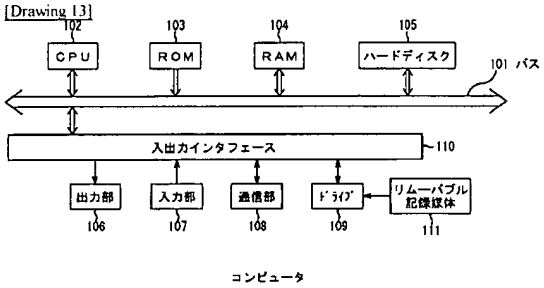


音声合成部 3

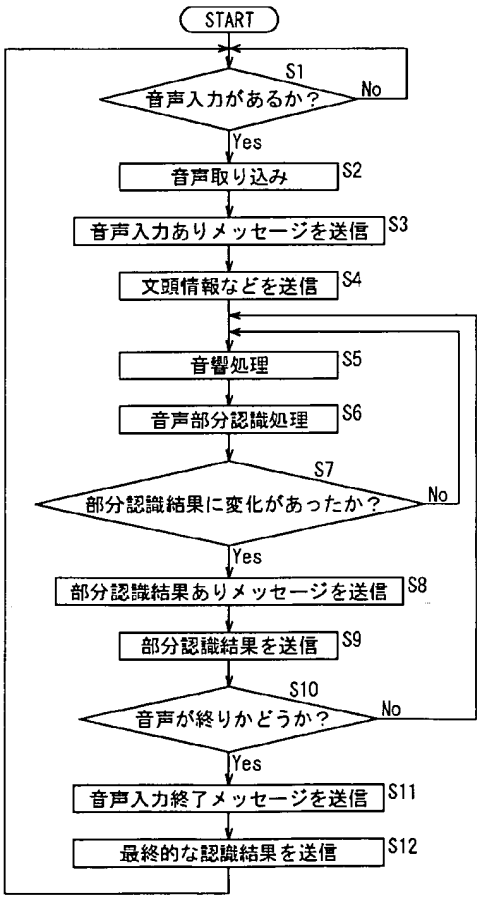
[Drawing 7]



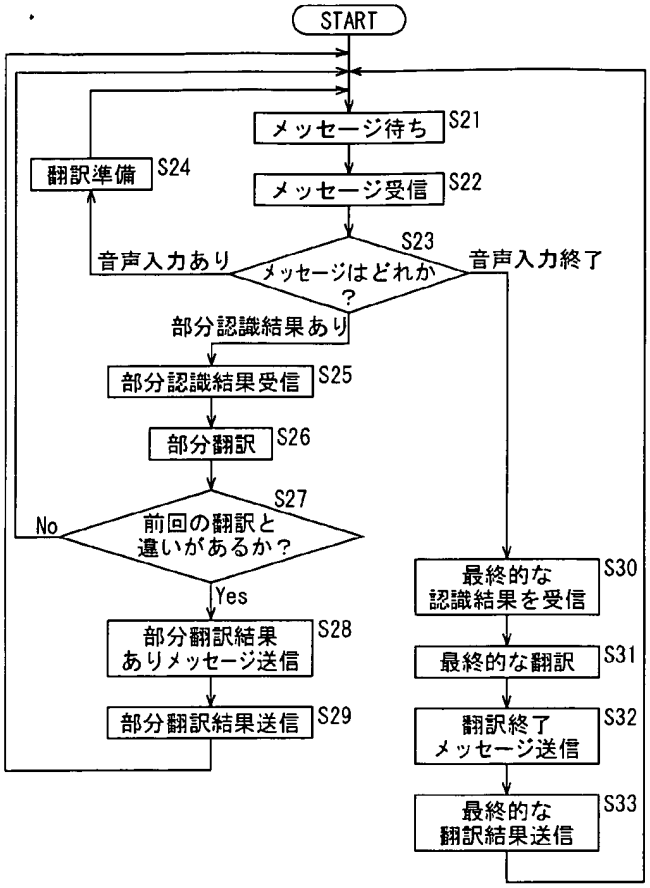
[Drawing 13]



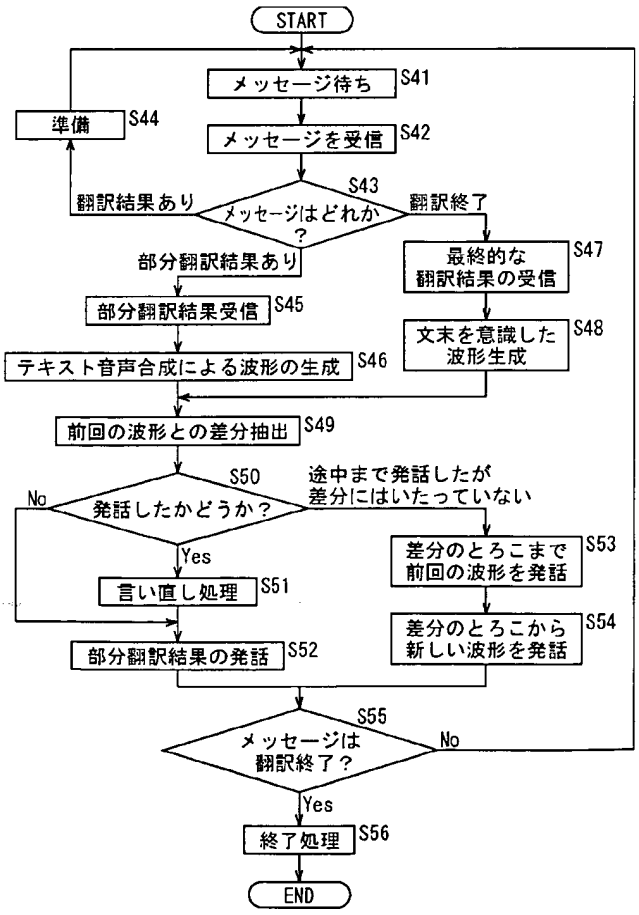
[Drawing 8]



[Drawing 9]


START
メッセージ待ち
S21
翻訳準備
S24
メッセージ受信
S22
音声入力あり
S23
メッセージはどれか？
音声入力終了
部分認識結果あり
部分認識結果受信
S25
部分翻訳
S26
No
S27
前回の翻訳と違いがあるか？
Yes
部分翻訳結果ありメッセージ送信
S28
部分翻訳結果送信
S29
最終的な認識結果を受信
S30
最終的な翻訳
S31
翻訳終了メッセージ送信
S32
最終的な翻訳結果送信
S33

[Drawing 10]


START
メッセージ待ち
S41
準備
S44
メッセージを受信
S42
翻訳結果あり
S43
メッセージはどれか？
翻訳終了
部分翻訳結果あり
最終的な翻訳結果の受信
S47
部分翻訳結果受信
S45
文末を意識した波形生成
S48
テキスト音声合成による波形の生成
S46
前回の波形との差分抽出
S49
No
S50
発話したかどうか？
途中まで発話したが差分にはいたっていない
Yes
言い直し処理
S51
差分のところまで前回の波形を発話
S53
差分のところから新しい波形を発話
S54
部分翻訳結果の発話
S52
S55
メッセージは翻訳終了？
No
Yes
終了処理
S56
END

[Drawing 11]

語順が大きく異なる［ので］、	Because of largely different word order,
↓	↓
語順が大きく異なる［ので］、 語順を保った［まま］	Because of largely different word order, with keeping the word order,
↓	↓
語順が大きく異なる［ので］、 語順を保った［まま］ 日本語を英語に翻訳する［のは］	Because of largely different word order, with keeping the word order, to translate Japanese into English is
↓	↓
語順が大きく異なる［ので］、 語順を保った［まま］ 日本語を英語に翻訳する［のは］ 難しい［と］	Because of largely different word order, with keeping the word order, to translate Japanese into English is difficult,
↓	↓
語順が大きく異なる［ので］、 語順を保った［まま］ 日本語を英語に翻訳する［のは］ 難しい［と］ 言われています。	Because of largely different word order, with keeping the word order, to translate Japanese into English is difficult, they say.
(A)	(B)

[Drawing 12]

英語と日本語とでは語順が大きく異なる［ので］、	Because English and Japanese are largely different in word order,
↓	↓
英語と日本語とでは語順が大きく異なる［ので］、 語順を保った［まま］	Because English and Japanese are largely different in word order, with keeping the word order,
↓	↓
英語と日本語とでは語順が大きく異なる［ので］、 語順を保った［まま］ 訳す［のは］	Because English and Japanese are largely different in word order, with keeping the word order, translation is
↓	↓
英語と日本語とでは語順が大きく異なる［ので］、 語順を保ったまま訳す［のは］	Because English and Japanese are largely different in word order, translation with keeping the word order is
↓	↓
英語と日本語とでは語順が大きく異なる［ので］、 語順を保ったまま訳す［のは］ 難しい［と］	Because English and Japanese are largely different in word order, translation with keeping the word order is difficult,
↓	↓
英語と日本語とでは語順が大きく異なる［ので］、 語順を保ったまま訳す［のは］ 難しい［と］ 言われています。	Because English and Japanese are largely different in word order, with keeping the word order, translation with keeping the word order is difficult, they say.
(A)	(B)

[Translation done.]